

WHAT IS CLAIMED IS:

1. A computer readable medium including instructions readable by a computer, which when implemented, cause the computer to classify data comprising the steps of:

receiving labeled data;  
receiving unlabeled data;  
constructing a first classifier and a second classifier using the labeled data;  
selecting a first uncertain portion from the unlabeled data that is uncertain with respect to the first classifier; and  
labeling the first uncertain portion using the second classifier to form a first labeled set.

2. The computer readable medium of claim 1, wherein receiving labeled data includes receiving data assigned to classes and wherein receiving unlabeled data includes receiving data capable of being assigned to classes.

3. The computer readable medium of claim 2, and further comprising reconstructing the first and second classifiers using at least the first labeled set.

4. The computer readable medium of claim 1, wherein selecting a first uncertain portion includes selecting instances of the unlabeled data as a function of uncertainty.

5. The computer readable medium of claim 4, wherein selecting instances as a function of uncertainty includes calculating probabilities.

6. The computer readable medium of claim 5, wherein calculating probabilities includes calculating probabilities that the first classifier is unable to label some instances of the unlabeled data.

7. The computer readable medium of claim 4, and further comprising calculating at least one value of uncertainty used to select the first uncertain portion.

8. The computer readable medium of claim 7, wherein selecting a first uncertain portion includes selecting instances having uncertainty values relative to a predetermined threshold.

9. The computer readable medium of claim 8, wherein selecting instances includes selecting instances having uncertainty values below the predetermined threshold.

10. The computer readable medium of claim 1, and further comprising:

selecting a second uncertain portion from the unlabeled data that is uncertain with respect to the second classifier; and

labeling the second uncertain portion using the first classifier to form a second labeled set.

11. The computer readable medium of claim 10, and further comprising exchanging information between the first labeled and the second labeled set to form at least one shared set.

12. The computer readable medium of claim 11, and further comprising reconstructing the first and second classifiers using the at least one shared set.

13. The computer readable medium of claim 10, wherein labeling the first uncertain portion includes assigning instances in the first labeled set to a first set of classes, and wherein labeling the second uncertain portion includes assigning instances in the second labeled set to a second set of classes.

14. The computer readable medium of claim 13, wherein the first set of classes and the second set of classes are the same.

15. The computer readable medium of claim 13, wherein the first set of classes and the second set of classes are different.

16. The computer readable medium of claim 13, wherein the first set of classes and the second set of classes are mutually exclusive.

17. The computer readable medium of claim 13, wherein the first set of classes and the second set of classes overlap by having classes in common.

18. The computer readable medium of claim 1, and further comprising selecting a first certain portion from the unlabeled data that is relatively certain with respect to the first classifier.

19. The computer readable medium of claim 18, wherein selecting a first certain portion includes selecting the most certain unlabeled data with respect to the first classifier.

20. The computer readable medium of claim 18, wherein selecting a first uncertain portion includes selecting data in the first certain portion.

21. The computer readable medium of claim 18, wherein selecting a first uncertain portion includes selecting data not in the first certain portion.

22. A computer readable medium including instructions readable by a computer, which when implemented, cause the computer to classify data comprising the steps of:

- constructing a first classifier and a second classifier using received labeled instances;
- selecting unlabeled instances that are certain with respect to the first classifier and

uncertain with respect to the second classifier data to form a first set of unlabeled instances; and  
selecting unlabeled instances that are certain with respect to the second classifier and uncertain with respect to the first classifier to form a second set of unlabeled instances.

23. The computer readable medium of claim 22, and further comprising:

labeling the first set of unlabeled instances using the first classifier to form a first labeled set;

labeling the second set of unlabeled instances using the second classifier; and to form a second labeled set;

adding the first labeled set and the second labeled set to the received labeled instances to form an augmented set.

24. The computer readable medium of claim 23, and further reconstructing the first classifier and the second classifier using the augmented set.

25. The computer readable medium of claim 24, wherein reconstructing the first and the second classifiers includes iteratively reconstructing the first and the second classifiers for each class of unlabeled instances.

26. A data classification system comprising the computer readable medium of claim 1.

27. A method of training a classifier, the method comprising:

receiving labeled data;

receiving unlabeled data;

constructing a first classifier and a second classifier using the labeled data;

selecting some of the unlabeled data that is certain with respect to the first classifier and uncertain with respect to the second classifier to form a first set of unlabeled data; and

selecting some of the unlabeled data that is certain with respect to the second classifier and uncertain with respect to the first classifier to form a second set of unlabeled data.

28. The method of claim 27 wherein receiving unlabeled data includes receiving data capable of being assigned to classes, and wherein receiving labeled data includes receiving data assigned to classes.

29. The method of claim 28, and further comprising reconstructing the first and the second classifier for each class.

30. The method of claim 27, and further comprising:  
applying the first classifier to the first  
unlabeled set to form a first labeled set;  
and  
applying the second classifier to the second  
labeled set to form a second labeled set.

31. The method of claim 30, and further comprising  
augmenting the received labeled data with the first  
labeled set and the second labeled set to form an  
augmented set.

32. The method of claim 31, and further comprising  
using the augmented set of labeled data to retrain  
the first classifier and the second classifier to  
form a retrained first classifier and a retrained  
second classifier.

33. The method of claim 30, wherein applying the  
first classifier to the first unlabeled set includes  
calculating probabilities that the first classifier  
is unable to assign some unlabeled data to classes.

34. The method of claim 33, wherein calculating  
probabilities includes calculating values of  
uncertainty.

35. The method of claim 34, wherein calculating values of uncertainty includes calculating values of uncertainty relative to a predetermined threshold.

36. A method of assigning information into classes, the method comprising:

- receiving labeled data;
- receiving unlabeled data;
- constructing a first classifier and a second classifier with the received labeled data;
- selecting a portion of unlabeled data that is uncertain for the first classifier; and
- assigning classes to the portion of unlabeled data using the second classifier to form a first labeled set.

37. The method of claim 36, and further comprising:

- selecting a portion of unlabeled data that is uncertain for the second classifier; and
- assigning classes to the portion of unlabeled data uncertain for the second classifier using the first classifier to form a second labeled set.